Pattern Recognition 48 (2015) 4049–4056

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Motion-towards-each-other-based hand gesture initialization

Zhiquan Feng [a,b], Bo Yang [a,b,*,1], Na Lv [a,b], Tao Xu [a,b], Jianqin Yin [a,b], Xiuyang Zhao [a,b], Shichang Feng [a,b,c]

[a] School of Information Science and Engineering, University of Jinan, Jinan 250022, China
[b] Provincial Key Laboratory for Network Based Intelligent Computing, Jinan 250022, China
[c] School of Information Science and Engineering, Shandong University of Science and Technology, 266590, China

ARTICLE INFO

Article history:
Received 27 February 2015
Received in revised form
12 May 2015
Accepted 16 May 2015
Available online 29 May 2015

Keywords:
3D human hand gesture model
Motion-towards-each-other-based interaction
Gesture recognition

ABSTRACT

Initialization for a 3D hand model is significant in determining a 3D hand model that corresponds to a user's 3D hand pose in the initial frame. We propose a motion-towards-each-other-based human–computer cooperation algorithm with the aim to reduce the operational burden on the user and cause the computer's response to the user's operation to fall within acceptable scopes. The computer first recognizes the user's interactive intentions through the hand gesture sensing algorithm and at last cooperates with the user until they achieve the meeting point. The main contributions of this study are sensing the operator's interactive intentions and focusing on the motion-towards-each-other-based interactive cooperation mechanisms between a human and a computer to achieve low cognitive and operational burdens. The proposed approach can effectively reduce users' cognitive and operational burdens. The proposed initialization system is successfully applied to several application systems.

1. Introduction

The objective of 3D hand pose reconstruction is to reconstruct the initial 3D hand models in accordance with the users' poses in the initial frames of an online hand video, from which recursive hand trackers can work.

Although the initialization of the visual tracking system is critical in the performance of freehand tracking systems, not much is known about how the procedure is implemented. Most of the currently available algorithms assume that initialization is manually performed.

Initialization for a 3D hand model is also an issue. First, recovering a 3D hand structure from a single 2D hand image is challenging. Secondly, the human hand is a typical articulated and elastic object with high dimensionality; establishing a real 3D hand model from nearly unlimited hand poses is almost impossible. We address this problem by motion-towards-each-other-based cooperation between a human and a computer in this paper.

Similar to the applications of the human–computer interface in 3D application systems, such as drag-and-drop, 3D virtual assembly, and point device, freehand tracking requires a reliable initial pose in the first frame for an easier, more pleasurable and satisfactory experience.

2. Related work

LaGorce et al. [1] assumed that the hand is parallel to the image plane at initialization and that linear constraints are defined on the relative length of the parts in each finger. Furthermore, the researchers assumed that hand color is constant across the surface. A few systems are featured with automatic initialization. Schlattmann et al. [2] developed an algorithm that is able to extract all the information required to solve both gesture-recognition and pose-estimation problems from camera images in real time. The information was computed from the 3D binary voxel grid of the visual hull created from the segmented views of the cameras. Many cameras were employed, and an algorithm that is able to extract all the information required to solve gesture-recognition and pose-estimation problems from the camera images in real time was developed. Thus, hand tracking and initialization of the hand model from the lost frame were realized. According to Stenger et al. [3], a hand model is automatically initialized by searching the complete tree in the first frame of the sequence. de la Rivière et al. [4] addressed the initialization problem at the beginning of each session, where the user was asked to stretch out his fingers while presenting the back of his hand to the camera. They extracted peaks and valleys from the image. Linking each

* Corresponding author at: School of Information Science and Engineering, University of Jinan, Shandong, 250022, China. Tel.: +86 0531 82767929.
E-mail address: yangbo@ujn.edu.cn (B. Yang).
[1] Current address: #336 Nanxinzhuang West Road, Jinan, Shandong, 250022, China.

http://dx.doi.org/10.1016/j.patcog.2015.05.015
0031-3203/

finger tip, represented by a peak, to the middle of the two neighboring valleys makes the finger axis appear. The deviation angle from the vertical axis (the middle finger) can be utilized to approximately evaluate the pose of the hand. According to Causo et al. [5], the hand is at the initial pose when the palm is flat open and the fingers extend away from the palm. Lee et al. [6] initialized a 3D hand model through ellipsoid fitting on top of the segmented hand silhouette and by computing an estimated scale and orientation of the hand. They refined the coarse model by minimizing the distance map along the gradient of each joint offset. Causo [7] initialized his hand tracking system by setting the state vector to zero value at the initial step. Zero values assigned to the state parameters mean that the hand model is at its initial pose. The hand is at the initial pose when the palm is flat open and the fingers extend away from the palm.

Feng et al. [8] initialized the 3D pose and position of the freehand by modeling the cognitive behavior of the operators and interacting between the operator and the computer. However, this method does not determine the ways in which the human operator and the computer effectively cooperate, it does not demonstrate why and how behavioral models guide our algorithm to assign tasks for the initialization between the operator and computer, and it doses not determine when the meeting point between the operator and the computer occur. Afterwards they proposed a key factor-based pose initialization algorithm to improve the time cost [9]. However, the issues above mentioned are still exist.

In most situations, initialization is manually performed or requires users to manually label the initial positions in practice. For example, Schwarz et al. [10] performed initialization by measuring the limb lengths and creating an initial labeling of specific landmarks. Several researchers assumed that the initial position of the hand is known [11,12].

3. Motion-towards-each-other based hand gesture initialization

Hand pose initialization aims to reconstruct the initial 3D hand gestures based on the poses of the user in the first frames of an online hand video. This study proposes a motion-towards-each-other based hand gesture initialization approach. With Our Method (which is called OM algorithm in this paper), initialization is fulfilled by means of moving towards each other between the operators and the computer. Suppose the initial statuses of hand gesture and scene object (which is the 3D hand model in this paper) are $\boldsymbol{S}_{0,\ hg}$ and $\boldsymbol{S}_{0,\ so}$ respectively, the status at the meeting point is noted as $\boldsymbol{S}_1$ (see Fig. 1). OM algorithm is described with the formulae (1)–(6).

$$\boldsymbol{S}_1 = \boldsymbol{OPERATOR}(\boldsymbol{S}_{0,hg}) \tag{1}$$

$$\boldsymbol{S}_1 = \boldsymbol{COMPUTER}(\boldsymbol{S}_{0,so}) \tag{2}$$

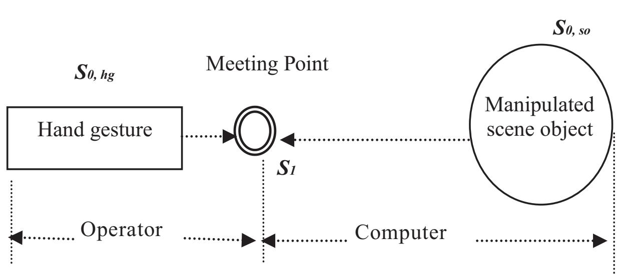

Fig. 1. Idea of the proposed OM algorithm.

$$\boldsymbol{OPERATOR} \otimes \boldsymbol{COMPUTER} < 0 \tag{3}$$

$$E(D_U) < T_1 \tag{4}$$

$$E(T_C) < T_2 \tag{5}$$

$$\text{Minimize}(\boldsymbol{\psi}(\boldsymbol{S}_1)). \tag{6}$$

The formulae (1) and (2) indicate that the operator exerts operation sequence $\boldsymbol{OPERATOR}$ on $\boldsymbol{S}_{0,\ hg}$ while the computer exerts operation sequence $\boldsymbol{COMPUTER}$ on $\boldsymbol{S}_{0,\ so}$, and by interactively cooperating with each other between the $\boldsymbol{OPERATOR}$ and the $\boldsymbol{COMPUTER}$, $\boldsymbol{S}_{0,\ hg}$ and $\boldsymbol{S}_{0,\ so}$ achieve the same status $\boldsymbol{S}_1$. The extinct characteristic of the cooperation is that the hand gesture and the scene object change towards each other by means of interactive cooperation between the two operation series which is depicted with formula (3)—the operational directions of $\boldsymbol{COMPUTER}$ and $\boldsymbol{OPERATOR}$ are contrasting. With pose transformation as an example, if the pose moves towards the left, the computer should make the interacted object move towards the right; if the pose moves towards the bottom, the computer should make the interacted object move towards the top. If the pose image becomes big (when the pose moves towards the camera), the computer should make the interacted object become small. In formula (4), $E(D_U)$ refers to the mathematical expectation of the averaged change in the user's hand gesture during the time interval from $\boldsymbol{S}_{0,\ hg}$ to $\boldsymbol{S}_1$. In formula (5), $E(T_C)$ refers to the mathematical expectation of the averaged time cost for the computer during the time interval from $\boldsymbol{S}_{0,\ so}$ to $\boldsymbol{S}_1$; T_1 and T_2 are thresholds. Therefore, Formulae (4) and (5) reduce the operational burden on the user and cause the computer's response to the user's operation to fall within acceptable scopes. Lastly, the meeting point between the two sequences is determined with the optimization function (6).

4. Optimization function

The objective of designing optimization function is to find the meeting point minimizing the operational burden on the user and the operational burden on the computer.

Suppose that L algorithms, M_1, M_2, …, M_L, are applied to the same initialization task. T_a is the average time cost of the L algorithms, A_k and T_k are the accuracy and time cost of the kth algorithm (where $1 \le k \le L$), and α_k is the degree of harmoniousness (DOH) determined by the users, which reflects the users' cognitive burden.

Cognitive burden was evaluated by four factors in this study: tiredness, joviality, freedom, and workability. Tiredness describes the extent of toil the user felt in the process of initialization. Joviality describes the degree of amusement the user felt. Convenience describes the quality of being suitable for the user's purposes, and workability describes the extent to which the initialization approach is feasible. The four factors were scored between 0 and 100 by users.

Our evaluation criterion is defined as Eq. (7).

$$\psi_k = \alpha_k T_a \frac{A_k}{T_k} \tag{7}$$

where α_k is the average of tiredness, joviality, convenience, and workability of the algorithm k.

The commonly used method to evaluate 3D hand models is to project the hand model onto the input image to show how well the projection matches the image data, which can be measured by Hausdorff distance.

We define accuracy as

$$A_k = e^{-\beta H_k}, \tag{8}$$